METHOD AND SYSTEM FOR DETECTION OF SUSPICIOUS LESIONS IN DIGITAL MAMMOGRAMS USING A COMBINATION OF SPICULATION AND DENSITY SIGNALS

Jimmy R. Roehrig
Takeshi Doi
Nico Karssemeijer
Guido M. te Brake

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application Serial No. 09/103,290, entitled "Method and Apparatus for Fast Detection of Spiculated Lesions in Digital Mammograms," filed June 23, 1998, which is a continuation of U.S. Patent Application Serial No. 08/676,660, filed July 10, 1996, ~~for which a notice of allowance was mailed on May 12, 1998 for which an issue fee was paid on May 19, 1998~~ Issued Sept 29, 1998 now Patent No. 5,915,591 issued 9/29/98. Both of the above applications are assigned to the assignee of the present invention. The above applications are hereby incorporated by reference into the present application.

FIELD OF THE INVENTION

The present invention relates to the field of computer aided diagnosis of medical images. In particular, the invention relates to a method and apparatus for detecting suspicious lesions in digital mammograms using an algorithm that independently computes mass information and spiculation information for allowing faster and more reliable identification of suspicious lesions.

BACKGROUND OF THE INVENTION

Breast cancer in women is a serious health problem. The American Cancer Society currently estimates that over 180,000 U.S. women are diagnosed with breast cancer each year. Breast cancer is the second major cause of cancer death among women. The American Cancer Society also estimates that breast cancer causes the death of over 44,000 U.S. women each year. While, at present, there is no means for preventing

breast cancer, early detection of the disease prolongs life expectancy and decreases the likelihood of the need for a total mastectomy. Mammography using x-rays is currently the most common method of detecting and analyzing breast tumors.

5   Detection of suspicious, i.e. possibly cancerous, areas in mammograms is an important first step in the early diagnosis and treatment of breast cancer. While it is important to detect suspicious lesions when they are in the early stages, practical considerations can make this

10 difficult. One complicating factor is that a typical mammogram may contain myriads of lines corresponding to fibrous breast tissue. The trained, focused eye of a medical professional, such as a radiologist, is needed to detect suspicious features among these lines. A typical radiologist

15 may be required to examine hundreds of mammograms per day, leading to the possibility of a missed diagnosis due to fatigue and human error.

Recently, medical professionals have begun to use Computer Aided Diagnostic (CAD) tools to assist them in

20 detecting suspicious features. Experiments have shown that the performance of radiologists improves when they are assisted by detection software that marks suspicious areas. *See* Brake and Karssemeijer, "Detection of Stellate Breast Abnormalities," Digital Mammography pp. 341-346 (Elsevier

25 Science 1996), the contents of which are hereby incorporated by reference into the present disclosure.

FIG. 1 shows a continuum of lesions that may appear in mammograms, ranging from a pure mass or pure density lesion on the left to a spiculated lesion on the right. Other

30 types of lesions include architectural distortions, which have radiating lines similar to spiculated lesions but are generally without a central mass, and radial scars, which appear as criss-crossed lines and also are generally without a central mass.

35   Sharply defined masses such as those shown at the left in FIG. 1 are rarely associated with malignant tumors, while spiculated masses can be a strong indication of malignancy.

3

Lesions having the characteristics of architectural distortions or radial scars may also be cancerous, depending on their size and shape.

Accordingly, there is value in locating and analyzing

5 both the "mass" or "density" qualities and the "spiculatedness" qualities of shapes found in digital mammograms. CAD systems generally include "mass" (or density) focused algorithms and "spiculation" focused algorithms. Some algorithms attempt to use metrics of both

10 "massness" and "spiculatedness" to identify suspicious portions of digital mammograms. Such algorithms are included among the following references: Yin and Giger et. al., "Computerized Detection of Masses in Digital Mammograms: Analysis of Bilateral Subtraction Images," Medical Physics,

15 Vol. 18, No. 5, pp. 955-963 (Sept./Oct. 1991); Sahiner et. al., "Classification of Masses on Mammograms Using a Rubber-Band Straightening Transform and Feature Analysis," Medical Imaging, SPIE Symposium on Medical Imaging Paper No. 2710-06, at p. 204 (1996); and Huo and Giger et al., "Analysis of

20 Spiculation in the Computerized Classification of Mammographic Masses," Medical Physics, Vol. 22, No. 10, pp. 1569-1579 (Oct. 1995). The contents of the above references are hereby incorporated by reference into the present application.

25 Typical of the above algorithms, Huo and Giger take a serial and dependent approach by first identifying masses and subsequently identifying spiculatedness characteristics of those masses. Huo and Giger demonstrate how the edges of detected masses can be used to determine a measure of

30 spiculation -- i.e., spiculatedness. Huo and Giger, approach the problem serially by first detecting the mass signature, or density, in a mammogram, and then applying various filtering analyses to filter out benign masses and false detections, such as parenchymal structure. The Huo and Giger

35 approach initially depends on density, which is a feature with low positive predictive value. It then attempts to improve upon itself by measuring features with higher

4

positive predictive values, such as spiculation. The
shortcomings of the Huo/Giger approach include the fact that
the additional feature measurements typically depend on
secondary algorithms that may be non-robust. These secondary

5   algorithms may include algorithms for spiculation, region
growing, or segmentation of the mass boundaries. Also, these
algorithms may display poor sensitivity on architectural
distortions and radial scars which have no central density.
Finally, because the algorithm is inherently serial, wherein

10 the spiculation information is computed after the mass
information, the time for completion of the algorithm is the
sum of the time for completion of the mass detection
algorithm plus the spiculation detection algorithm, which can
lead to disadvantageously slow results.

15   A direct "backward direction" algorithm for of
spiculation detection is disclosed in Karssemeijer,
"Recognition of Stellate Lesions in Digital Mammograms,"
Digital Mammography: Proceedings of the 2nd International
Workshop on Digital Mammography, York, England, 10-12 July,

20 1994, pp. 211-219 (Elsevier Science 1994), the contents of
which are hereby incorporated by reference into the present
application. By "backward direction" it is meant that a
"candidate point" is incrementally moved across the image by
a distance corresponding to the desired resolution of the

25 spiculation search. At each candidate point, a set of
"window computations" for a window of pixels surrounding the
candidate point is performed, and a metric corresponding to
the presence and/or strength of a spiculation centered on the
candidate point is computed.

30   "Backward direction" algorithms are computationally
intensive. For an image size of N x N pixels, there
generally need to be on the order of $K(bN)^2$ computations,
where K is the number of window computations for each
candidate point and b is the reciprocal of the number of

35 image pixels between each candidate point. Because the
number K is often proportional to the square or cube of the
window size, the computational intensity of "backward

- 4 -   PEMP-105312.1

5

direction" approaches can easily become unwieldy. Another
example of a "backward direction" spiculation algorithm is
described in Kegelmeyer, "Computer-aided Mammographic
Screening for Spiculated Lesions," Radiology, Vol. 191, pp.
5 331-337 (1994), the contents of which are hereby
incorporated by reference into the present application. The
computational complexity of "backward direction" spiculation
algorithms may cause a CAD program to be too slow for
practical use by medical professionals, such as radiologists.
10 Additionally, a practical implementation of a CAD system
using a backward direction algorithm for spiculation
detection would lead to inevitable dependency between the
mass and spiculation algorithms. This is because, due to its
slowness, the spiculation algorithm could only be applied to
15 a subset of interesting regions of the digital mammogram
image, the interesting regions being pointed out by the
presence of masses from the mass detection algorithm.

Accordingly, it would be desirable to provide a
computer-assisted diagnosis (CAD) system for assisting in the
20 detection of suspicious lesions in medical images that has
increased speed in computing the necessary mass information
and spiculation information.

It would be further desirable to provide a computer-
assisted diagnosis (CAD) system that has greater reliability
25 in detecting suspicious lesions of a digital mammogram that
have characteristics similar to those of architectural
distortions, radial scars, and spiculated lesions having very
small central masses.

30

35

PEMP-105312.1

SUMMARY OF THE INVENTION

These and other objects of the present invention are achieved by an improved CAD system that incorporates

5   independent measurements of mass and spiculation characteristics as taken from a digital mammogram image. In a preferred embodiment, a forward spiculation detection algorithm is incorporated that is executed separately from a mass detection algorithm, and that is not dependent on any

10 results from the mass detection algorithm. The step of computing spiculation information may be performed before, during, or after the step of computing mass information, but in a preferred embodiment is performed concurrently in time with the mass detection algorithm for optimal time

15 performance of the overall system.

According to a preferred embodiment, after the mass information and spiculation information is determined, a classifier algorithm classifies locations in the digital mammogram according to their feature vectors, which comprise

20 the computed mass information and spiculation information. The classifier algorithm may incorporate any of a variety of classification algorithms known in the art, including linear classifier algorithms, quadratic classifier algorithms, K-nearest-neighbor classifier algorithms, decision tree

25 classifier algorithms, or neural network classifier algorithms. By way of non-limiting example, the classifier algorithm may classify locations into a two-category system including a "suspicious" category and a "normal" category.

30

35

PEMP-105312.1

7

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects, features and advantages of the present invention will be more readily apparent from the following detailed description of the invention in which:

5   FIG. 1 shows various types of lesions arranged in an order from pure mass to highly spiculated mass;

FIG. 2A is an outside view of an illustrative computer aided diagnostic (CAD) system;

FIG. 2B is a block diagram of an illustrative CAD

10 processing unit for use in the CAD system of FIG. 2A;

FIG. 3 is a flowchart showing steps taken by a CAD system in accordance with a preferred embodiment;

FIG. 4 is a flowchart showing further steps taken by a CAD system in accordance with a preferred embodiment.

15

DETAILED DESCRIPTION OF THE INVENTION

FIG. 2A is an outside view of an illustrative computer aided diagnostic (CAD) system 100 for assisting in the identification of suspicious areas in mammograms according to

20 the preferred embodiment. CAD system 100 comprises a CAD processing unit 102 and a viewing station 104. In general, CAD processing unit 102 scans a developed x-ray mammogram 101 into a digital mammogram image, processes the image, and outputs a highlighted digital mammogram for viewing at

25 viewing station 104.

FIG. 2B is a block diagram of CAD processing unit 102. In accordance with the invention, processing unit 102 is capable of performing a multiplicity of image processing algorithms designed to detect abnormalities such as

30 spiculation detection, mass detection, linear weighted comparisons and general mathematical comparisons, either serially or in parallel with the disclosed abnormality detection algorithms. Preferably, CAD processing unit 102 includes a digitizer 103, a central control unit 105, a

35 memory 108, a parallel processing unit 110, and I/O unit 112. Digitizer 103 illustratively is a scanner with 50 micron resolution.

Viewing station 104 is for conveniently viewing both the x-ray mammogram 101 on a backlighting station 120 and the output of the CAD processing unit 102 on a display device 118. The display device 118 may be, for example, a CRT

5   screen. The display device 118 typically shows a highlighted digital mammogram corresponding to the x-ray mammogram 101, the highlighted digital mammogram having information directing the attention of the radiologist to suspicious areas which may contain spiculation as determined by image

10 processing steps performed by the CAD processing unit 102. In one embodiment of the invention, the highlighted digital mammogram will have black or red circles circumscribing locations with suspected abnormalities. Since the x-ray mammogram 101 on backlighting station 120 and the digitized

15 mammogram on display device 118 are physically adjacent one another, one application of viewing station 104 is to use the digitized mammogram to direct the attention of the radiologist to the spiculated portions of the actual x-ray mammogram 101 itself.

20   It is to be appreciated that the CAD processing unit 102 is capable of performing other image processing algorithms on the digital mammogram in addition to or in parallel with the algorithms for detecting abnormalities in accordance with the preferred embodiment. In this manner, the radiologist may be

25 informed of several suspicious areas of the mammogram at once by viewing the display device 118, spiculation being one special type of suspicious area.

        After the x-ray mammogram 101 passes through the CAD system 100, it undergoes processing similar to that currently

30 practiced in clinics. In addition, memory 108 of CAD processing unit 102 may be used in conjunction with I/O unit 112 to generate a permanent record of the highlighted digital mammogram described above, and/or may also be used to allow non-real-time viewing of the highlighted digital mammogram.

35   FIG. 3 is an overview showing steps performed by CAD processing unit 102 on the x-ray mammogram in accordance with a preferred embodiment. At step 302, the x-ray mammogram is

scanned in and digitized into a digital mammogram. The
digital mammogram may be, for example, a 3000 x 4000 array of
12-bit gray scale pixel values. Such a digital mammogram
would generally correspond to a typical 8" x 10" x-ray

5   mammogram which has been digitized at a 50 micron spatial
resolution. Because a full resolution image such as the 3000
x 4000 image described above is not necessary for the
effectiveness of the preferred embodiment, the image may be
locally averaged, using steps known in the art, down to a

10 smaller size corresponding, for example, to a 200 micron
spatial resolution. At such a resolution, a typical image
would then be an M x N array of 12-bit gray scale pixel
values, with M being near 900, for example, and N being near
1200, for example. In general, however, either the full

15 resolution image or the locally averaged image may be used as
the original digital mammogram in accordance with the
preferred embodiment.

        At step 304, a spiculation detection algorithm is
performed on the digital mammogram. At step 306 a mass

20 detection algorithm is run on the digital mammogram. In a
preferred embodiment, steps 304 and 306 are carried out
concurrently so as to optimize overall speed of the detection
process. Also in a preferred embodiment, steps 304 and 306
are carried out independently, in that there is no data

25 dependence between them. According to a preferred
embodiment, the spiculation detection step 304 does not
require any final or intermediate outputs from the mass
detection step 306, and the mass detection step 306 does not
require any final or intermediate outputs from the

30 spiculation detection step 304. In addition to introducing
the ability to make the overall algorithm faster, the
independence of the mass detection and spiculation detection
steps allows for increased detection of features
characteristic of architectural distortions, radial scars,

35 and in general otherwise suspicious lesions that do not have
a significant central mass that is detected by the mass
detection algorithm.

PEMP-105312.1

/O

At step 308, a classifier algorithm is performed on feature vectors corresponding to locations in the digital mammogram. Each location in the digital mammogram has a corresponding feature vector, which can be defined as a set

5   of characteristics, including the "massness" metrics and "spiculatedness" metrics that were determined previously at steps 304 and 306. A classifier is an algorithm or system that labels a feature vector as belonging to a specific class, such as "suspicious/normal," "malignant/benign," etc.

10 Several types of classifiers exist in the art, including linear classifiers, quadratic classifiers, k-nearest-neighbor method classifiers, decision trees, and neural networks. As known in the art, classifiers are constructed using a data set of example vectors representing each class, called a

15 training set or learning set. *See generally* Brake & Karssemeijer, *supra*, and references cited therein. Finally, at step 312, the digital mammogram image and a list of suspicious lesions are sent to the viewing station 104 for display.

20   Many spiculation detection algorithms are known in the art. Any number of these spiculation detection algorithms may be used to accomplish step 304. For example, step 304 may be satisfied by running either a backward or a forward spiculation detection algorithm. However, backward

25 spiculation detection algorithms consume substantially more time, computer memory, and general resources than forward spiculation detection algorithms. For this reason, a preferred embodiment uses a forward spiculation detection algorithm. The information resulting from a spiculation

30 detection algorithm may include such details as the geometrical coordinates of mammographic areas likely to contain spiculations, the size of the spiculation, or the like.

        In step 306 of FIG. 3, a mass detection algorithm is run

35 against the digitized mammogram. Importantly, step 306 is independent of step 304 in that the mass detection algorithm is run against the digitized mammogram without reference to

- 10 -

//

the spiculation detection algorithm of step 304, or the results of the spiculation detection algorithm of step 304. The independent nature of steps 304 and 306 is important because it solves the problem of that prior art in which a

5 mass detection algorithm and a spiculation detection algorithm are serially applied. In the prior art, a mass detection algorithm is applied to the digitized image first so that areas with "density," or mass, can be identified. In this way, a mass detection algorithm making a "first cut"

10 singles out only areas that may contain suspicious masses. Next, a spiculation detection algorithm is applied only to those suspicious areas that may contain masses, rather than to the entire mammographic image. Thus, spiculated masses with low density may be overlooked.

15   The preferred embodiment solves the "first cut" problem of the prior art because both a spiculation detection algorithm, step 304, and a mass detection algorithm, step 306, are independently applied to the entire mammographic image. By applying both a spiculation detection algorithm

20 and a mass detection algorithm to the entire image, an increased number of suspicious areas are likely to be identified. For example, areas containing low density spiculated masses may be identified. Step 306 can be executed using any number of mass detection algorithms known

25 in the art.

        FIG. 4 shows exemplary steps corresponding to the classification step 308 when a simple linear classifier method is used. The example of FIG. 4 is presented for clarity and completeness of disclosure, to allow the reader

30 to more fully comprehend the context of the preferred embodiment, and is not intended to limit the scope of the present invention. It is to be understood that any of a variety of classifiers can be used at step 308, each having certain advantages, disadvantages, and tradeoffs in terms of

35 training time, computation time, probability of false positives, probability of missed detection, and other factors. The method described herein has an advantage,

- 11 -   PEMP-105312.1

12

however, of simplicity and speed of training and computation time.

The classification method of FIG. 4 supposes the common result that the spiculation detection algorithm and the mass detection algorithm produce a scalar spiculation metric f_s and a scalar mass metric f_m, respectively, each having a value normalized between 0 and 100, for example. Thus, for example, a given location may have a scalar spiculation metric f_s of 90, which would indicate a very high degree of spiculatedness, and a scalar mass metric f_m of 5, which would indicate a very faint or small mass. In this simple case, the pair (f_s, f_m) forms the feature vector at each location.

At step 402, the scalar spiculation metric f_s and the scalar mass metric f_m are each multiplied by weighting factors a_s and a_m, respectively, and the weighted factors are added to form a result. At step 404, the result from step 402 is compared to a predetermined threshold C. If the result from step 402 is greater than the predetermined threshold C, the location is identified as "suspicious" at step 406. Otherwise, the location is identified as "normal" at step 408. The choices for a_s, a_m, and C define the parameters for the linear classifier algorithm of FIG. 4. Using methods known in the art, these linear classifier parameters are statistically preselected using a large training set of feature vectors that are known to represent each class being identified.

As known in the art, the parameters of the classifier may be changed to achieve different levels of utility and results. Even in the simple linear classifier example of FIG. 4, it is readily observed that the predetermined threshold C may be lowered to increase system sensitivity and bring more mammograms to the attention of the radiologist. This would, of course, have the negative impact of increasing the overall number of mammograms and false positives that the radiologist must analyze. Conversely, the predetermined threshold C may be increased to decrease system sensitivity, and would have the converse result of necessitating less

PEMP-105312.1

13

radiologist intervention while risking more missed diagnoses.

The linear classifier technique of FIG. 4 can be
generalized and extended to include several feature vector
5 metrics (f1, f2, f3, ...) and several weights (a1, a2, a3 . .
.). Other feature vector metrics may include, for example,
the "sphericity" and "eccentricity" metrics disclosed in the
parent U.S. App. No. 09/103,290, *supra*.

It is also within the scope of the preferred embodiments
10 to for the classifier to identify more than two classes. For
example, a classifier may be constructed to form four sets --
"suspicious -- more spiculated"; "suspicious -- more
density;" "suspicious -- similar density and spiculation",
and "normal." One use of the multiple classes would be, for
15 example, to place red triangles around "suspicious -- more
density" locations on the CRT display 118, placing blue
triangles around "suspicious – more spiculation" locations,
and green squares around "suspicious -- similar density and
spiculation" locations. In this manner, the radiologist
20 would be made aware of different types of suspicious lesions
in different ways.

An alternative embodiment uses a look up table to
combine the independent information at each location in the
digitized mammogram. The look up table might be a two
25 dimensional matrix that indicates for numerical values of
mass information on a first axis and spiculation information
on a second axis whether the combination of mass information
and spiculation information is suspicious.

Both linear classifiers and neural networks are types of
30 classifiers. In order to be able to use a classifier, prior
statistical knowledge about the digital mammograms is
necessary. For example, see Brake and Karssemeijer,
"Detection of Stellate Breast Abnormalities," Digital
Mammography pp. 341-346 (Elsevier Science 1996). Other types
35 of classifiers include the Bayes optimal classifier,
quadratic classifiers, the *kth*-nearest neighbor classifier,
and artificial neural networks, or the like. One skilled in

- 13 -

PEMP-105312.1

|4

the art will know that any of these pattern classification systems could be used according to the preferred embodiments.

Once an abnormality is located, the preferred embodiment could be extended to advantageously create further

5   classification parameters to indicate areas according to probable seriousness or likelihood of locating a benign abnormality versus locating a malignant abnormality. Such a further classification step could assist the medical professional in decision making and prioritization.

10   Although preferred embodiments have been described with respect to a CAD system for detecting suspicious lesions in digital mammograms, those skilled in the art should be able to apply the preferred embodiments to any number of other computer aided diagnosis systems.

15

20

25

30

35

PEMP-105312.1

15